

k -AUTOMATIC SETS OF RATIONAL NUMBERS

ERIC ROWLAND AND JEFFREY SHALLIT

ABSTRACT. The notion of a k -automatic set of integers is well-studied. We develop a new notion — the k -automatic set of rational numbers — and prove basic properties of these sets, including closure properties and decidability.

1. INTRODUCTION

Let k be an integer ≥ 2 , and let $\mathbb{N} = \{0, 1, 2, \dots\}$ denote the set of non-negative integers. Let $\Sigma_k = \{0, 1, \dots, k-1\}$ be an alphabet. Given a word $w = a_1 a_2 \cdots a_t \in \Sigma_k^*$, we let $[w]_k$ denote the integer that it represents in base k ; namely,

$$(1) \quad [w]_k = \sum_{1 \leq i \leq t} a_i k^{t-i},$$

where (as usual) an empty sum is equal to 0. For example, $[101011]_2 = 43$.

Note that in this framework, every element of \mathbb{N} has infinitely many distinct representations as words, each having a certain number of *leading zeroes*. Among all such representations, the one with no leading zeroes is called the *canonical representation*; it is an element of $C_k := \{\epsilon\} \cup (\Sigma_k - \{0\})\Sigma_k^*$. For an integer $n \geq 0$, we let $(n)_k$ denote its canonical representation.

Given a language $L \subseteq \Sigma_k^*$, we can define the set of integers it represents, as follows:

$$(2) \quad [L]_k = \{[w]_k : w \in L\}.$$

Definition 1. We say that a set $S \subseteq \mathbb{N}$ is k -automatic if there exists a regular language $L \subseteq \Sigma_k^*$ such that $S = [L]_k$.

The class of k -automatic sets of natural numbers has been widely studied (e.g., [3, 4, 1]), and many properties of these sets are known. For example, it is possible to state an equivalent definition involving only canonical representations:

Definition 2. A set $S \subseteq \mathbb{N}$ is k -automatic if the language $(S)_k := \{(n)_k : n \in S\}$ is regular.

To see the equivalence of Definitions 1 and 2, note that if L is a regular language, then so is the language L' obtained by removing all leading zeroes from each word in L .

A slightly more general concept is that of k -automatic *sequence*. Let Δ be a finite alphabet. Then a sequence (or infinite word) $(a_n)_{n \geq 0}$ over Δ is said to be k -automatic if, for every $c \in \Delta$, the set of *fibers* $F_c = \{n \in \mathbb{N} : a_n = c\}$ is a k -automatic set of natural numbers. Again, this class of sequences has been widely studied [3, 4, 1]. The following result is well-known [4]:

Theorem 1. The sequence $(a_n)_{n \geq 0}$ is k -automatic iff its k -kernel, the set of its subsequences $K = \{(a_{k^e n + f})_{n \geq 0} : e \geq 0, 0 \leq f < k^e\}$, is finite.

In a previous paper [11], the second author extended the notion of k -automatic sets over \mathbb{N} to sets over $\mathbb{Q}^{\geq 0}$, the set of non-negative rational numbers. In this paper, we will obtain some basic results about this new class. Our principal results are Theorem 4 (characterizing the sets of integers), Corollary 2 (showing that k -automatic sets of rationals are not closed under intersection), and Theorem 9 (showing that it is decidable if a k -automatic set is infinite).

Our class has some similarity to another class studied by Even [5] and Hartmanis and Stearns [6]; their class corresponds to the topological closure of our k -automatic sets, where the denominators are restricted to powers of k .

2. REPRESENTING RATIONAL NUMBERS

A natural representation for the non-negative rational number p/q is the pair (p, q) with $q \neq 0$. Of course, this representation has the drawback that every element of $\mathbb{Q}^{\geq 0}$ has infinitely many representations, each of the form $(jp/d, jq/d)$ for some $j \geq 1$, where $d = \gcd(p, q)$.

We might try to ensure uniqueness of representations by considering only “reduced” representations (those in “lowest terms”), which amounts to requiring that $\gcd(p, q) = 1$. However, this condition cannot be checked by automata — see Remark 1 below — and deciding if an arbitrary regular language consists entirely of reduced representations is not evidently computable. Furthermore, insisting on reduced representations rules out the representation of some reasonable sets of rationals, such as $\{(k^m - 1)/(k^n - 1) : m, n \geq 1\}$ (see Theorem 6). For these reasons, **we allow the rational number p/q to be represented by *any* pair of integers (p', q') with $p/q = p'/q'$.**

Next, we need to see how to represent a pair of integers as a word over a finite alphabet. Here, we follow the ideas of Salon [8, 9, 10]. Consider the alphabet Σ_k^2 . A finite word w over Σ_k^2 can be considered as a sequence of pairs $w = [a_1, b_1][a_2, b_2] \cdots [a_n, b_n]$ where $a_i, b_i \in \Sigma_k$ for $1 \leq i \leq n$. We can now define the projection maps π_1, π_2 , as follows:

$$\pi_1(w) = a_1 a_2 \cdots a_n; \quad \pi_2(w) = b_1 b_2 \cdots b_n.$$

Note that the domain and range of each projection are $(\Sigma_k^2)^*$ and Σ_k^* , respectively. Then we define $[w]_k = ([\pi_1(w)]_k, [\pi_2(w)]_k)$. Thus, for example, if $w = [1, 0][0, 1][1, 0][0, 0][1, 1][1, 0]$, then $[w]_2 = (43, 18)$.

In this framework, every pair of integers (p, q) again has infinitely many distinct representations, arising from padding on the left by leading pairs of zeroes, that is, by $[0, 0]$. Among all such representations, the *canonical representation* is the one having no leading pairs of zeroes. We write it as $(p, q)_k$.

We now state the fundamental definitions of this paper:

Definition 3. *Given a word $w \in (\Sigma_k^2)^*$ with $[\pi_2(w)]_k \neq 0$, we define*

$$\text{quo}_k(w) := \frac{[\pi_1(w)]_k}{[\pi_2(w)]_k}.$$

If $[\pi_2(w)]_k = 0$, then $\text{quo}_k(w)$ is not defined. Further, if $[\pi_2(w)]_k \neq 0$ for all $w \in L$, then $\text{quo}_k(L) := \{\text{quo}_k(w) : w \in L\}$. A set of rational numbers $S \subseteq \mathbb{Q}^{\geq 0}$ is k -automatic if there exists a regular language $L \subseteq (\Sigma_k^2)^$ such that $\text{quo}_k(L) = S$.*

Note that in our definition, a given rational number α can have multiple representations in two different ways: it can have non-canonical representations that

begin with leading zeroes, and it can have “unreduced” representations (p, q) where $\gcd(p, q) > 1$.

Given a set $S \subseteq \mathbb{Q}$, if S contains a non-integer, then by calling it k -automatic, it is clear that we intend this to mean automatic in the sense of this section. But what if $S \subseteq \mathbb{N}$? Then calling it “automatic” might mean automatic in the usual sense, as in Section 1, or in the extended sense introduced in this section, treating S as a subset of \mathbb{Q} . In Theorem 4 we will see that these two interpretations actually *coincide* for subsets of \mathbb{N} , but in order to prove this, we need some notation to distinguish between the two types of representations. So, by (\mathbb{N}, k) -automatic we mean the interpretation in Section 1 and by (\mathbb{Q}, k) -automatic we mean the interpretation in this section.

So far we have only considered representations where the leftmost digit is the most significant digit. However, sometimes it is simpler to deal with *reversed representations* where the leftmost digit is the least significant digit. In other words, sometimes it is easier to deal with the reversed word w^R and reversed language L^R instead of w and L , respectively. Since the regular languages are (effectively) closed under reversal, for most of our results it does not matter which representation we choose, and we omit extended discussion of this point.

We use the following notation for intervals: $I[\alpha, \beta]$ denotes the closed interval $\{x : \alpha \leq x \leq \beta\}$, and similarly for open- and half-open intervals.

3. EXAMPLES

To build intuition, we give some examples of k -automatic sets of rationals.

Example 1. Let $k = 2$, and consider the regular language L_1 defined by the regular expression $A^*\{[0, 1], [1, 1]\}A^*$, where $A = \{[0, 0], [0, 1], [1, 0], [1, 1]\}$. This regular expression specifies all pairs of integers where the second component has at least one nonzero digit — the point being to avoid division by 0. Then $\text{quo}_k(L) = \mathbb{Q}^{\geq 0}$, the set of all non-negative rational numbers.

Example 2. Consider the regular language

$$L_2 = \{w \in (\Sigma_k^2)^* : \pi_1(w) \in C_k \cup \{0\} \text{ and } [\pi_2(w)]_k = 1\}.$$

Then $\text{quo}_k(L_2) = \mathbb{N}$.

Example 3. Let $k = 3$, and consider the regular language L_3 defined by the regular expression $[0, 1]\{[0, 0], [2, 0]\}^*$. Then $\text{quo}_k(L_2)$ is the *3-adic Cantor set*, the set of all rational numbers in the “middle-thirds” Cantor set with denominators a power of 3 [2].

Example 4. Let $k = 4$, and consider the set $S = \{0, 1, 3, 4, 5, 11, 12, 13, \dots\}$ of all non-negative integers that can be represented using only the digits 0, 1, -1 in base 4. Consider the set $L_5 = \{(p, q)_4 : p, q \in S\}$. It is not hard to see that L_5 is $(\mathbb{Q}, 4)$ -automatic. The main result in [7] can be phrased as follows: $\text{quo}_4(L_5)$ contains every odd integer. In fact, an integer t is in $\text{quo}_4(L_5)$ if and only if the exponent of the largest power of 2 dividing t is even.

Example 5. Note that $\text{quo}_k(L_1 \cup L_2) = \text{quo}_k(L_1) \cup \text{quo}_k(L_2)$ but the analogous identity involving intersection need not hold. As an example, consider $L_1 = \{[2, 1]\}$ and $L_2 = \{[4, 2]\}$. Then $\text{quo}_{10}(L_1 \cap L_2) = \emptyset \neq \{2\} = \text{quo}_{10}(L_1) \cap \text{quo}_{10}(L_2)$.

4. BASIC RESULTS

In this section we obtain some basic results about automatic sets of rationals. First we state a result from [11]:

Theorem 2. *Let β be a non-negative real number and define*

$$L_{\leq\beta} = \{w \in (\Sigma_k^2)^* : \text{quo}_k(w) \leq \beta\},$$

and analogously for the other relations $<, =, \geq, >, \neq$. Then $L_{\leq\beta}$ (resp., $L_{<\beta}, L_{=\beta}, L_{\geq\beta}, L_{>\beta}$) is regular iff β is a rational number.

Lemma 1. *Let M be a DFA with input alphabet Σ_k^2 and let $F \subseteq \mathbb{Q}^{\geq 0}$ be a finite set of integers. Then the following problems are recursively solvable:*

- (1) *Is $F \subseteq \text{quo}_k(L(M))$?*
- (2) *Is $\text{quo}_k(L(M)) \subseteq F$?*

Proof. To decide if $F \subseteq \text{quo}_k(L(M))$, we simply check, using Theorem 2, whether $x \in \text{quo}_k(L(M))$ for each $x \in F$.

To decide if $\text{quo}_k(L(M)) \subseteq F$, we create DFA's accepting $L_{=x}$ for each $x \in F$, using Theorem 2. Now we create a DFA accepting all representations of all elements of F using the usual "direct product" construction for the union of regular languages. Since F is finite, the resulting DFA A is finite. Now, using the usual direct product construction, we create a DFA accepting $L(A) - L(M)$ and check to see if its language is nonempty. ■ □

Suppose S is a set of real numbers, and α is a real number. We introduce the following notation:

$$\begin{aligned} S + \alpha &:= \{x + \alpha : x \in S\} & S \dot{-} \alpha &:= \{\max(x - \alpha, 0) : x \in S\} \\ \alpha \dot{-} S &:= \{\max(\alpha - x, 0) : x \in S\} & \alpha S &:= \{\alpha x : x \in S\}. \end{aligned}$$

Theorem 3. *The class of k -automatic sets of rational numbers is closed under the following operations: (i) union; (ii) $S \rightarrow S + \alpha$ for $\alpha \in \mathbb{Q}^{\geq 0}$; (iii) $S \rightarrow S \dot{-} \alpha$ for $\alpha \in \mathbb{Q}^{\geq 0}$; (iv) $S \rightarrow \alpha \dot{-} S$ for $\alpha \in \mathbb{Q}^{\geq 0}$; (v) $S \rightarrow \alpha S$ for $\alpha \in \mathbb{Q}^{\geq 0}$; (vi) $S \rightarrow \{1/x : x \in S \setminus \{0\}\}$.*

Proof. We prove only item (2), with the others being similar. We will use the reversed representation, with the least significant digit appearing first. Write $\alpha = p/q$. Let M be a DFA with $\text{quo}_k(L(M)) = S$. To accept $S + \alpha$, on input a base- k representation of $x = p'/q'$, we transduce the numerator to $p'q - pq'$ and the denominator to qq' (hence effectively computing a representation for $x - \alpha$), and simultaneously simulate M on this input, digit-by-digit, accepting if M accepts. ■ □

We now state one of our main results.

Theorem 4. *Let $S \subseteq \mathbb{N}$. Then S is (\mathbb{N}, k) -automatic if and only if it is (\mathbb{Q}, k) -automatic.*

The proof requires a number of preliminary results. First, we introduce some terminology and notation. We say a set $S \subseteq \mathbb{N}$ is *ultimately periodic* if there exist integers $n_0 \geq 0, p \geq 1$ such that $n \in S \iff n + p \in S$, provided $n \geq n_0$. In particular, every finite set is ultimately periodic. We let $\mathcal{P} = \{2, 3, 5, \dots\}$ denote the set of prime numbers. Given a positive integer n , we let $\text{pd}(n)$ denote the set

of its prime divisors. For example, $\text{pd}(60) = \{2, 3, 5\}$. Given a subset $D \subset \mathcal{P}$, we let $\pi(D) = \{n \geq 1 : \text{pd}(n) \subseteq D\}$, the set of all integers that can be factored completely using only the primes in D . Finally, let $k \geq 2$, $n \geq 1$ be integers, and define $\nu_k(n) := \max\{i : k^i \mid n\}$, the exponent of the largest power of k dividing n .

The first result we need is the following theorem, which is of independent interest.

Theorem 5. *Let $D \subseteq \mathcal{P}$ be a finite set of prime numbers, and let $S \subseteq \pi(D)$. Then S is k -automatic iff*

1. $F := \{\frac{s}{k^{\nu_k(s)}} : s \in S\}$ is finite

and

2. For all $f \in F$ the set $U_f = \{i : k^i f \in S\}$ is ultimately periodic.

First, we need a lemma.

Lemma 2. *Let D be a finite set of prime numbers, and let $S \subseteq \pi(D)$. Let s_1, s_2, \dots be an infinite sequence of (not necessarily distinct) elements of S . Then there is an infinite increasing sequence of indices $i_1 < i_2 < \dots$ such that $s_{i_1} \mid s_{i_2} \mid \dots$.*

Proof. Case 1: The sequence (s_i) is bounded. In this case infinitely many of the s_i are the same, so we can take the indices to correspond to these s_i .

Case 2: The sequence (s_i) is unbounded. In this case we prove the result by induction on $|D|$. If $|D| = 1$, then we can choose a strictly increasing subsequence of the (s_i) ; since all are powers of some prime p , this subsequence has the desired property.

Now suppose the result is true for all sets D of cardinality $t - 1$. We prove it for $|D| = t$. Since only t distinct primes figure in the factorization of the s_i , some prime, say p , must appear with unbounded exponent in the (s_i) . So there is some subsequence of (s_i) , say (t_i) , with strictly increasing exponents of p . Now consider the infinite sequence (u_i) given by $u_i = t_i / p^{\nu_p(t_i)}$. Each u_i has a prime factorization in terms of the primes in $D - \{p\}$, so by induction there is an infinite increasing sequence of indices i_1, i_2, \dots such that $u_{i_1} \mid u_{i_2} \mid \dots$. Then $p^{\nu_p(t_{i_1})} u_{i_1} \mid p^{\nu_p(t_{i_2})} u_{i_2} \mid \dots$, which corresponds to an infinite increasing sequence of indices of the original sequence (s_i) . ■

Now we can prove Theorem 5.

Proof. \Leftarrow : S can be written as the disjoint finite union $\bigcup_{f \in F} f \cdot \{k^i : i \in U_f\}$, where U_f is an ultimately periodic set of integers. Each term in the union has base- k representation $(f)_k \{0^i : i \in U_f\}$ and hence is regular. It follows that S is k -automatic.

\Rightarrow : Suppose F is infinite. Notice that, from the definition, no element of F is divisible by k . Therefore we can write S as the disjoint union $\bigcup_{i \geq 0} k^i H_i$, where $H_i := \{f \in F : k^i f \in S\}$. Then there are two possibilities: either (a) the sets H_i are finite for all $i \geq 0$, or (b) at least one H_i is infinite.

In case (a), define $u_i := \max H_i$ for all $i \geq 0$. Then the set $\{u_0, u_1, \dots\}$ must be infinite, for otherwise F would be finite. Choose an infinite subsequence of the u_i consisting of distinct elements, and apply Lemma 2. Then there is an infinite increasing subsequence of indices $i_1 < i_2 < \dots$ such that $u_{i_1} \mid u_{i_2} \mid \dots$. So the sequence $(u_{i_j})_{j \geq 1}$ is strictly increasing.

Now consider the characteristic sequence of S , say $(f(n))_{n \geq 0}$, taking values 1 if $n \in S$ and 0 otherwise. Consider the subsequences (f_j) in the k -kernel of f defined

by $f_j(n) = f(k^{i_j}n)$ for $n \geq 0$, $j \geq 1$. By our construction, the largest n in $\pi(D)$ such that $k \nmid n$ and $f_j(n) = 1$ is $n = u_{i_j}$. Since the u_{i_j} are strictly increasing, this shows the (infinitely many) sequences (f_j) are pairwise distinct. Hence, by Theorem 1, f is not k -automatic and neither is S .

In case (b), we have H_i is infinite for some i . As mentioned above, S can be written as the disjoint union $\bigcup_{i \geq 0} k^i H_i$. Let L be the language of canonical base- k expansions of elements of H_i (so that, in particular, no element of L starts with 0). The base- k representation of elements of $k^i H_i$ end in exactly i 0's, and no other member of S has this property. Since S is assumed to be k -automatic, it follows that L is regular. Note that no two elements of H_i have a quotient which is divisible by k , because if they did, the numerator would be divisible by k , which is ruled out by the condition.

Since L is infinite and regular, by the pumping lemma, there must be words u, v, w , with v nonempty, such that $uv^jw \in L$ for all $j \geq 0$. Note that for all integers $j \geq 0$ and $c \geq 0$ we have

$$(3) \quad [uv^{j+c}w]_k = [uv^jw]_k \cdot k^{c|v|} + ([v^c w]_k - [w]_k \cdot k^{c|v|}).$$

Let $D = \{p_1, p_2, \dots, p_t\}$. Since $[uv^jw]_k \in S \subseteq \pi(D)$, it follows that there exists a double sequence $(f_{r,j})_{1 \leq r \leq t; j \geq 1}$ of non-negative integers such that

$$(4) \quad [uv^jw]_k = p_1^{f_{1,j}} \dots p_t^{f_{t,j}}$$

for all $j \geq 0$. From (4), we see that $k^{|uw|+j|v|} < p_1^{f_{1,j}} \dots p_t^{f_{t,j}}$, and hence (assuming $p_1 < p_2 < \dots < p_t$) we get

$$(|uw| + j|v|) \log k < \left(\sum_{1 \leq r \leq t} f_{r,j} \right) \log p_t.$$

Therefore, there are constants $0 < c_1$ and J such that $c_1 j < \sum_{1 \leq r \leq t} f_{r,j}$ for $j \geq J$.

For each $j \geq J$ now consider the indices r such that $f_{r,j} > c_1 j/t$; there must be at least one such index, for otherwise $f_{r,j} \leq c_1 j/t$ for each r and hence $\sum_{1 \leq r \leq t} f_{r,j} \leq c_1 j$, a contradiction. Now consider $t+1$ consecutive j 's; for each j there is an index r with $f_{r,j} > c_1 j/t$, and since there are only t possible indices, there must be an r and two integers l' and l , with $0 \leq l < l' \leq t$, with $f_{r,j+l} > c_1(j+l)/t$ and $f_{r,j+l'} > c_1(j+l')/t$. This is true in any block of $t+1$ consecutive j 's that are $\geq J$. Now there are infinitely many disjoint blocks of $t+1$ consecutive j 's, and so there must be a single r and a single difference $l' - l$ that occurs infinitely often. Put $\delta = l' - l$.

Now use (3) and take $c = \delta$. We get infinitely many j such that

$$p_1^{f_{1,j+\delta}} \dots p_t^{f_{t,j+\delta}} = k^{\delta|v|} p_1^{f_{1,j}} \dots p_t^{f_{t,j}} + E,$$

where $E = [v^\delta w]_k - [w]_k \cdot k^{\delta|v|}$ is a constant that is independent of j . Now focus attention on the exponent of p_r on both sides. On the left it is $f_{r,j+\delta}$, which we know to be at least $c_1(j+\delta)/t$. On the right the exponent of p_r dividing the first term is $f_{r,j} + \delta|v|e_r$ (where $k = p_1^{e_1} \dots p_t^{e_t}$); this is at least $c_1 j/t$. So p_r^h divides E , where $h \geq c_1 j/t$. But this quantity goes to ∞ , and E is a constant. So $E = 0$. But then

$$\frac{[uv^{j+\delta}w]_k}{[uv^jw]_k} = k^{\delta|v|}.$$

which is impossible, since, as we observed above, two elements of H_i cannot have a quotient which is a power of k . This contradiction shows that H_i cannot be infinite.

So now we know that F is finite. Fix some $f \in F$ and consider $T_f = \{k^i : k^i f \in S\}$. Since S is k -automatic, and the set of base- k expansions $(T_f)_k$ essentially is formed by stripping off the bits corresponding to $(f)_k$ from the front of each element of S of which $(f)_k$ is a prefix, and replacing it with “1”, this is just a left quotient followed by concatenation, and hence $(T_f)_k$ is regular. Let M' be a DFA for $(T_f)_k$, and consider an input of 1 followed by l 0's for $l = 0, 1, \dots$ in M' . Evidently we eventually get into a cycle, so this says that $U_f = \{i : k^i f \in S\}$ is ultimately periodic.

This completes the proof of Theorem 5. ■

□

Corollary 1. *Suppose S is a k -automatic set of integers accepted by a finite automaton M . There is an algorithm to decide, given M , whether there exists a finite set $D \subseteq \mathcal{P}$ such that $S \subseteq \pi(D)$. Furthermore, if such a D exists, we can also determine the sets F and U_f in Theorem 5.*

Proof. To determine if such a D exists, it suffices to remove all trailing zeroes from words in $(S)_k$ and see if the resulting language is finite. If it is, we know F , and then it is a simple matter to compute the U_f . ■

□

We can now prove Theorem 4.

Proof. One direction is easy, since if S is (\mathbb{N}, k) -automatic, then there is an automaton accepting $(S)_k$. We can now easily modify this automaton to accept all words over $(\Sigma_k^2)^*$ whose π_2 projection represents the integer 1 and whose π_1 projection is an element of $(S)_k$. Hence S is (\mathbb{Q}, k) -automatic.

Now assume $S \subseteq \mathbb{N}$ is (\mathbb{Q}, k) -automatic. If S is finite, then the result is clear, so assume S is infinite. Let L be a regular language with $\text{quo}_k(L) = S$. Without loss of generality we may assume every representation in L is canonical; there are no leading $[0, 0]$'s. Furthermore, by first intersecting with $L_{\neq 0}$ we may assume that L contains no representations of the integer 0. Finally, we can also assume, without loss of generality, that no representation contains *trailing* occurrences of $[0, 0]$, for removing all trailing zeroes from all words in L preserves regularity, and it does not change the set of numbers represented, as it has the effect of dividing the numerator and denominator by the same power of k . Since the words in L represent integers only, the denominator of every representation must divide the numerator, and hence if the denominator is divisible by k , the numerator must also be so divisible. Hence removing trailing zeroes *also* ensures that no denominator is divisible by k . Let M be a DFA of n states accepting L .

We first show that the set of possible denominators represented by L is actually finite. Write $S = S_1 \cup S_2$, where $S_1 = S \cap I[0, k^{n+1})$ and $S_2 = S \cap I[k^{n+1}, \infty)$. Let $L_1 = L \cap L_{<k^{n+1}}$, the representations of all numbers $< k^{n+1}$ in L , and $L_2 = L \cap L_{\geq k^{n+1}}$. Both L_1 and L_2 are regular, by Theorem 2. It now suffices to show that S_2 is (\mathbb{N}, k) -automatic.

Consider any $t \in S_2$. Let $z \in L_2$ be a representation of t . Since $t \geq k^{n+1}$, clearly $|z| \geq n$, and so $\pi_2(z)$ must begin with at least n 0's. Then, by the pumping lemma, we can write $z = uvw$ with $|uv| \leq n$ and $|v| \geq 1$ such that $uv^i w \in L$ for all $i \geq 0$. However, by the previous remark about $\pi_2(z)$, we see that $\pi_2(v) = 0^j$ for $1 \leq j \leq n$. Hence $[\pi_2(z)]_k = [\pi_2(uvw)]_k = [\pi_2(uw)]_k$. Since uw must also represent a member of S , it must be an integer, and hence $[\pi_2(z)]_k \mid [\pi_1(uw)]_k$ as well as

$[\pi_2(z)]_k \mid [\pi_1(uvw)]_k$. Hence

$$[\pi_2(z)]_k \mid [\pi_1(uvw)]_k - [\pi_1(uw)]_k = ([\pi_1(uw)]_k - [\pi_1(u)]_k) \cdot k^{|w|}.$$

The previous reasoning applies to any $z \in L_2$. Furthermore, $0 < [\pi_1(uw)]_k - [\pi_1(u)]_k < k^n$. It follows that every possible denominator d of elements in L_2 can be expressed as $d = d_1 \cdot d_2$, where $1 \leq d_1 < k^n$ and $d_2 \mid k^m$ for some m . It follows that the set of primes dividing all denominators d is finite, and we can therefore apply Theorem 5. Since k divides no denominator, the set of possible denominators is finite.

We can therefore decompose L_2 into a finite disjoint union corresponding to each possible denominator d . Next, we use a finite-state transducer to divide the numerator and denominator of the corresponding representations by d . For each d , this gives a new regular language A_d where the denominator is 1. Writing $T := \bigcup_d A_d$, we have $S_2 = \text{quo}_k(T) = \bigcup_d \text{quo}_k(A_d)$. Now we project, throwing away the second coordinate of elements of T ; the result is regular and hence S is a k -automatic set of integers. ■ □

As corollaries, we get that the k -automatic sets of rationals are (in contrast with sets of integers) not necessarily closed under the operations of intersection and complement.

Theorem 6. *Let $S_1 = \{(k^n - 1)/(k^m - 1) : 1 \leq m < n\}$ and $S_2 = \mathbb{N}$. Then $S_1 \cap S_2$ is not a k -automatic set.*

Proof. We can write every element of S_1 as p/q , where $p = (k^n - 1)/(k - 1)$ and $q = (k^m - 1)/(k - 1)$. The base- k representation of p is 1^n and the base- k representation of q is 1^m . Thus a representation for S_1 is given by the regular expression $[1, 0]^+ [1, 1]^+$. We know that \mathbb{N} is k -automatic from Example 2.

From a classical result we know that $(k^m - 1) \mid (k^n - 1)$ if and only if $m \mid n$. It follows that $S_1 \cap S_2 = T$, where $T = \{(k^n - 1)/(k^m - 1) : 1 < m < n \text{ and } m \mid n\}$. If the k -automatic sets of rationals were closed under intersection, then T would be (\mathbb{N}, k) -automatic. Writing $n = md$, we have $(k^n - 1)/(k^m - 1) = k^{(d-1)m} + \dots + k^m + 1$, whose base- k representation is $(10^{m-1})^{d-1}1$. Hence $(T)_k = \{(10^{m-1})^{d-1}1 : m \geq 1, d > 1\}$. Assume this is regular. Intersecting with the regular language $10^*10^*10^*1$ we get $\{10^n10^n10^n1 : n \geq 1\}$. But a routine argument using the pumping lemma shows this is not even context-free, a contradiction. ■ □

Corollary 2. *The class of (\mathbb{Q}, k) -automatic sets is not closed under the operations of intersection or complement.*

Proof. We have just shown that this class is not closed under intersection. But since it is closed under union, if it were closed under complement, too, it would be closed under intersection, a contradiction. ■ □

Remark 1. The technique above also allows us to prove that the languages $L_d = \{(p, q)_k : q \mid p\}$, $L_r = \{(p, q)_k : \gcd(p, q) > 1\}$, and $L_g = \{(p, q)_k : \gcd(p, q) = 1\}$ are not context-free.

We can also prove a decidability result.

Theorem 7. *Given a regular language $L \subseteq (\Sigma_k^2)^*$, it is decidable whether*

- (a) $\text{quo}_k(L) \subseteq \mathbb{N}$; (b) $\text{quo}_k(L) \setminus \mathbb{N}$ is finite.

Proof. For (a), given M accepting L , we first create (using Theorem 2) a DFA M_1 accepting the set $T_1 := (\text{quo}_k(L) \cap I[0, k^{n+1})) \setminus \{0, 1, \dots, k^{n+1} - 1\}$. If $T_1 \neq \emptyset$, then answer “no” and stop. Next, create a DFA M_2 accepting the set $T_2 := (\text{quo}_k(L) \cap I[k^{n+1}, \infty))$. Now project onto the second coordinate, and, using Corollary 1, decide if the integers so represented are factorable into a finite set of primes. If not, answer “no” and stop. Finally, using Corollary 1 again, create the decomposition in that corollary (resulting in only finitely many denominators) and check for each denominator whether it divides all of its corresponding numerators, using a transducer.

For (b) we need to check instead whether T_1 is a finite set, and whether there are finitely many exceptions to the second and third steps. ■ □

5. SOLVABILITY RESULTS

In this section we show that a number of problems involving k -automatic sets of rational numbers are recursively solvable. We start with a useful lemma.

Lemma 3. *Let $u, v, w \in (\Sigma_k^2)^*$ such that $|v| \geq 1$, and such that $[\pi_1(uvw)]_k$ and $[\pi_2(uvw)]_k$ are not both 0. Define*

$$(5) \quad U := \begin{cases} \text{quo}_k(w), & \text{if } [\pi_1(uv)]_k = [\pi_2(uv)]_k = 0; \\ \infty, & \text{if } [\pi_1(uv)]_k > 0 \text{ and } [\pi_2(uv)]_k = 0; \\ \frac{[\pi_1(uv)]_k - [\pi_1(u)]_k}{[\pi_2(uv)]_k - [\pi_2(u)]_k}, & \text{otherwise.} \end{cases}$$

(a) *Then exactly one of the following cases occurs:*

- (i) $\text{quo}_k(uw) < \text{quo}_k(uvw) < \text{quo}_k(uv^2w) < \dots < U$;
- (ii) $\text{quo}_k(uw) = \text{quo}_k(uvw) = \text{quo}_k(uv^2w) = \dots = U$;
- (iii) $\text{quo}_k(uw) > \text{quo}_k(uvw) > \text{quo}_k(uv^2w) > \dots > U$.

(b) *Furthermore, $\lim_{i \rightarrow \infty} \text{quo}_k(uv^i w) = U$.*

Proof. Part (a) is proved in [11]. We now prove the assertion about the limit.

Let $j \in \{1, 2\}$, let i be an integer ≥ 1 , and consider the base- k representation of the rational number

$$\frac{[\pi_j(uv^i w)]_k}{k^{i|v|+|w|}}; \quad \text{it looks like} \quad \pi_j(u). \overbrace{\pi_j(v)\pi_j(v) \cdots \pi_j(v)}^i.$$

On the other hand, the base- k representation of

$$[\pi_j(u)]_k + \frac{[\pi_j(v)]_k}{k^{|v|} - 1} \quad \text{looks like} \quad \pi_j(u). \pi_j(v)\pi_j(v) \cdots .$$

Subtracting, we get

$$\left| \frac{[\pi_j(uv^i w)]_k}{k^{i|v|+|w|}} - \left([\pi_j(u)]_k + \frac{[\pi_j(v)]_k}{k^{|v|} - 1} \right) \right| < k^{-i|v|}.$$

It follows that

$$\lim_{i \rightarrow \infty} \frac{[\pi_j(uv^i w)]_k}{k^{i|v|+|w|}} = [\pi_j(u)]_k + \frac{[\pi_j(v)]_k}{k^{|v|} - 1}.$$

Furthermore, this limit is 0 if and only if $[\pi_j(uv)]_k = 0$. Hence, provided $[\pi_2(uv)]_k \neq 0$, we get

$$\begin{aligned} \lim_{i \rightarrow \infty} \text{quo}_k(uv^i w) &= \lim_{i \rightarrow \infty} \frac{[\pi_1(uv^i w)]_k}{[\pi_2(uv^i w)]_k} = \lim_{i \rightarrow \infty} \frac{\frac{[\pi_1(uv^i w)]_k}{k^{i|v|+|w|}}}{\frac{[\pi_2(uv^i w)]_k}{k^{i|v|+|w|}}} = \frac{\lim_{i \rightarrow \infty} \frac{[\pi_1(uv^i w)]_k}{k^{i|v|+|w|}}}{\lim_{i \rightarrow \infty} \frac{[\pi_2(uv^i w)]_k}{k^{i|v|+|w|}}} \\ &= \frac{[\pi_1(u)]_k + \frac{[\pi_1(v)]_k}{k^{|v|-1}}}{[\pi_2(u)]_k + \frac{[\pi_2(v)]_k}{k^{|v|-1}}} = \frac{[\pi_1(uv)]_k - [\pi_1(u)]_k}{[\pi_2(uv)]_k - [\pi_2(u)]_k} = U. \end{aligned}$$

■

□

We now give some solvability results.

Theorem 8. *The following problems are recursively solvable: given a DFA M , a rational number α , and a relation \triangleleft chosen from $=, \neq, <, \leq, >, \geq$, does there exist $x \in \text{quo}_k(L(M))$ with $x \triangleleft \alpha$?*

Proof. The following gives a procedure for deciding if $x \triangleleft \alpha$. First, we create a DFA M' accepting the language $L_{\triangleleft \alpha}$ as described in Theorem 2 above. Next, using the usual direct product construction, we create a DFA M'' accepting $L(M) \cap L_{\triangleleft \alpha}$. Then, using breadth-first or depth-first search, we check to see whether there exists a path from the initial state of M'' to some final state of M'' . ■ □

Theorem 9. *The following problem is recursively solvable: given a DFA M , and an integer k , is $\text{quo}_k(L(M))$ of infinite cardinality?*

Note that this is *not* the same as asking whether $L(M)$ itself is of infinite cardinality, since a number may have infinitely many representations.

Proof. Without loss of generality, we may assume that the representations in M are canonical (contain no leading $[0, 0]$'s). Define

$$\gamma_k(u, v) = \frac{[\pi_1(uv)]_k - [\pi_1(u)]_k}{[\pi_2(uv)]_k - [\pi_2(u)]_k},$$

and let $\text{pref}(L)$ denote the language of all prefixes of all words of L . Let n be the number of states in M . We claim that $\text{quo}_k(L(M))$ is of finite cardinality if and only if $\text{quo}_k(L(M)) \subseteq T$, where

$$(6) \quad T = \{\text{quo}_k(x) : x \in L(M) \text{ and } |x| < n\} \cup \{\gamma_k(u, v) : uv \in \text{pref}(L) \text{ and } |v| \geq 1 \text{ and } |uv| \leq n\}.$$

One direction is easy, since if $\text{quo}_k(L(M)) \subseteq T$, then clearly $\text{quo}_k(L(M))$ is of finite cardinality, since T is.

Now suppose $\text{quo}_k(L(M)) \not\subseteq T$, so there exists some $x \in L(M)$ with $\text{quo}_k(x) \notin T$. Since T contains all words of $L(M)$ of length $< n$, such an x is of length $\geq n$. So the pumping lemma applies, and there exists a decomposition $x = uvw$ with $|uv| \leq n$ and $|v| \geq 1$ such that $uv^i w \in L$ for all $i \geq 0$. Now apply Lemma 3. If case (b) of that lemma applies, then $\text{quo}_k(x) = \gamma_k(u, v) \in T$, a contradiction. Hence either case (a) or case (c) must apply, and the lemma shows that $\text{quo}_k(uv^i w)$ for $i \geq 0$ gives infinitely many distinct elements of $\text{quo}_k(L(M))$.

To solve the decision problem, we can now simply enumerate the elements of T and use Lemma 1. ■ □

Theorem 10. *Given $p/q \in \mathbb{Q}^{\geq 0}$, and a DFA M accepting a k -automatic set of rationals S , it is decidable if p/q is an accumulation point of S .*

Proof. The number α is an accumulation point of a set of real numbers S if and only if at least one of the following two conditions holds:

$$(i) \quad \alpha = \sup(S \cap I(-\infty, \alpha)); \quad (ii) \quad \alpha = \inf(S \cap I(\alpha, \infty)).$$

By Theorem 2 we can compute a DFA accepting $S' := S \cap I(-\infty, \alpha)$ (resp., $S \cap I(\alpha, \infty)$). By [11, Thm. 2] we can compute $\sup S'$ (resp., $\inf S'$). ■ □

REFERENCES

- [1] Allouche, J.P., Shallit, J.: Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press (2003)
- [2] Cantor, G.: Über unendliche, lineare Punktmanigfaltigkeiten V. Math. Annalen 21, 545–591 (1883)
- [3] Cobham, A.: On the base-dependence of sets of numbers recognizable by finite automata. Math. Systems Theory 3, 186–192 (1969)
- [4] Cobham, A.: Uniform tag sequences. Math. Systems Theory 6, 164–192 (1972)
- [5] Even, S.: Rational numbers and regular events. IEEE Trans. Electr. Comput. 13, 740–741 (1964)
- [6] Hartmanis, J., Stearns, R.E.: Sets of numbers defined by finite automata. Amer. Math. Monthly 74, 539–542 (1967)
- [7] Loxton, J.H., van der Poorten, A.J.: An awful problem about integers in base four. Acta Arith 49, 193–203 (1987)
- [8] Salon, O.: Suites automatiques à multi-indices. Séminaire de Théorie des Nombres de Bordeaux 4(27), 4.01–4.27 (1986-1987)
- [9] Salon, O.: Suites automatiques à multi-indices et algébricité. C. R. Acad. Sci. Paris 305, 501–504 (1987)
- [10] Salon, O.: Propriétés arithmétiques des automates multidimensionnels. Ph.D. thesis, Université Bordeaux I (1989)
- [11] Shallit, J.: The critical exponent is computable for automatic sequences. In: Ambroz, P., Holub, S., Masáková, Z. (eds.) Proceedings 8th International Conference, WORDS 2011, pp. 231–239 (2011), vol. 63 of *Elect. Proc. Theor. Comput. Sci.*. Available at <http://arxiv.org/abs/1104.2303v2>